



Genome-wide analysis of homeobox genes from *Mesobuthus martensii* reveals Hox gene duplication in scorpions



Zhiyong Di ^{a,b}, Yao Yu ^c, Yingliang Wu ^a, Pei Hao ^d, Yawen He ^a, Huabin Zhao ^a, Yixue Li ^d, Guoping Zhao ^c, Xuan Li ^{d,*}, Wenxin Li ^{a,*}, Zhijian Cao ^{a,*}

^a College of Life Sciences, Wuhan University, Wuhan, China

^b College of Life Sciences, University of Science and Technology of China, Hefei, China

^c Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

^d Key Lab of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

ARTICLE INFO

Article history:

Received 25 March 2014

Received in revised form

10 April 2015

Accepted 12 April 2015

Available online 21 April 2015

Keywords:

Homeobox genes

Scorpion

Genome-wide analysis

Hox cluster

ABSTRACT

Homeobox genes belong to a large gene group, which encodes the famous DNA-binding homeodomain that plays a key role in development and cellular differentiation during embryogenesis in animals. Here, one hundred forty-nine homeobox genes were identified from the Asian scorpion, *Mesobuthus martensii* (Chelicerata: Arachnida: Scorpiones: Buthidae) based on our newly assembled genome sequence with approximately 248 × coverage. The identified homeobox genes were categorized into eight classes including 82 families: 67 ANTP class genes, 33 PRD genes, 11 LIM genes, five POU genes, six SINE genes, 14 TALE genes, five CUT genes, two ZF genes and six unclassified genes. Transcriptome data confirmed that more than half of the genes were expressed in adults. The homeobox gene diversity of the eight classes is similar to the previously analyzed Mandibulata arthropods. Interestingly, it is hypothesized that the scorpion *M. martensii* may have two Hox clusters. The first complete genome-wide analysis of homeobox genes in Chelicerata not only reveals the repertoire of scorpion, arachnid and chelicerate homeobox genes, but also shows some insights into the evolution of arthropod homeobox genes.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Homeobox genes were first discovered in the fruit fly (McGinnis et al., 1984; Scott and Weiner, 1984). They belong to a large gene group which includes 11 classes (Holland et al., 2007). Homeobox genes encode a highly conserved homeodomain. A homeobox is usually 180 base pairs (bp) long and encodes 60 amino acids that bind DNA (Bürglin, 1994, 2005). Homeobox genes play a key role in development and cellular differentiation during embryogenesis in simple and complex organisms.

The Hox cluster is a group of clustered homeobox genes, and also usually named the Hox genes that play important roles in pattern formation along the anterior-posterior body axis. They are assembled in two groups: the Bithorax complex (BX-C) and the

Antennapedia complex (ANT-C) in *Drosophila melanogaster* (see the review in Heffer and Pick (2013)).

Recently, an increasing number of genomes of different species have been sequenced, which provides the data for analyzing homeobox genes on a genome-wide level. Genome-wide analysis of homeobox genes has been completed for several animals and indicates that 255 (plus 78 pseudogenes or unassigned) homeobox genes were identified in human, 279 (plus 45 pseudogenes) in mouse, 133 in amphioxus, and 92 in nematode (<http://homeodb.zoo.ox.ac.uk/>).

In the phylum Arthropoda, the genome-wide analyses of the homeobox genes of several model insects have recently been completed. In the Mandibulata lineage of arthropods, the fruit fly *D. melanogaster* has 104, the beetle *Tribolium castaneum* has 105, and the honeybee *Apis mellifera* has 93 homeobox genes (<http://homeodb.zoo.ox.ac.uk/>). The silkworm *Bombyx mori* was reported to have 102 homeobox genes and was found to have a special group of 12 tandemly duplicated homeobox genes located between Bmpb and Bmzen, indicating that this Hox cluster had experienced a

* Corresponding authors. Tel.: +86 2768756746.

E-mail addresses: lixuan@sibs.ac.cn (X. Li), liwxlab@whu.edu.cn (W. Li), zjcao@whu.edu.cn (Z. Cao).

lineage-specific expansion in silkworms (Chai et al., 2008). Only two genome sequences have been determined in Chelicerates, *Tetranychus urticae* (2011) (Grbić et al., 2011) and *Ixodes scapularis* (2010, unpublished). Furthermore, a partial survey of homeobox genes was performed in *T. urticae*. The spider mite *T. urticae* was found to contain 8 of the canonical 10 Hox genes. Among them, the *ftz* gene is present in duplicate in two closely linked copies, but orthologues of Hox3 and abdominal A (*abdA*) were not found (Grbić et al., 2011). The morphological evolution of the segment loss in mites is speculated to correlate with the loss of a Hox gene (Grbić et al., 2011).

Known as “living fossils”, scorpions are one of the oldest arthropods. Scorpiones is a basal lineage of Arachnida, and Arachnida forms a major branch of Chelicerata. Their body is divided into three parts (tagmata): the head (prosoma), the abdomen (mesosoma) and the tail (metasoma). They have eight legs, two remarkable claws (pedipalps), particular pectines, and a narrow segmented tail ending with a venomous telson. The body morphology of mites consists of an anterior prosoma and posterior opisthosoma and is further distinguished by an extremely reduced body plan. The body morphology and structure of *T. urticae* is distinctly simpler than the scorpion. The unique morphological structures of scorpions imply that they can be used as a model for investigating the patterns, evolution and functions of homeobox and Hox genes. Recently, Sharma et al. (2014) reported that the scorpion *Centruroides sculpturatus* has two paralogues of all Hox genes except Hox3, suggesting cluster and/or whole genome duplication in this arachnid order.

In this study, we characterized homeobox genes from the *Mesobuthus martensii* genome. Transcriptome data revealed the expression profiles of the analyzed homeobox genes. In particular, the structure of the Hox cluster was analyzed in detail in *M. martensii*. *M. martensii* was hypothesized to have two Hox clusters (Hox A and Hox B). In total, we identified 149 homeobox genes, which were grouped into eight classes, including 82 families. This is the first complete genome-wide analysis of homeobox genes in Chelicerata. The results not only show that the Chelicerata lineage possibly has several patterns of Hox clusters, but also provide new insight into the structure and evolution of Hox cluster genes in arthropods.

2. Materials and methods

2.1. Identification and classification of homeobox genes

The new version of the scorpion *M. martensii* genome sequence and predicted protein database (version 1.0 gene models: <http://lifecenter.sgst.cn/main/en/Scorpion-Suppl/gene-models-v1.0.gff.zip>) was used to search for homeobox genes, as in the previously reported methods (Cao et al., 2013). The whole-genome sequencing project and genome assemblies of *M. martensii* are deposited in the GenBank database under BioProject PRJNA171479. The complete homeodomain protein sequences of 10 species, including human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), frog (*Xenopus (Silurana) tropicalis*), zebrafish (*Danio rerio*), amphioxus (*Branchiostoma floridae*), nematode (*Caenorhabditis elegans*), fruit fly (*D. melanogaster*), beetle (*T. castaneum*), and honeybee (*A. mellifera*) were downloaded from HomeoDB (Homeobox Database) (Supplementary file 1) (<http://homeodb.zoo.ox.ac.uk/>; Zhong and Holland, 2011) and the HGR (National Human Genome Research Institute) (NCBI). Homeodomain sequences from these ten organisms (HomeoDB) were collected and used as queries for BLASTP searches in the predicted protein database of the scorpion *M. martensii* genome. Considering that some homeodomain sequences cannot be classified because of low identity with the

reported homeobox family members, we analyzed all blast results with an identity value greater than 30%.

All of the retrieved candidate homeobox genes were further validated using a program in the SMART database to identify whether the protein sequences encoded by the candidate genes contain homeodomains (same method as Chai et al. (2008)). Candidate homeobox genes were further subjected to a detailed manual survey, particularly those with truncated genome annotations and atypical homeoboxes (NCBI, HomeoDB). The classification scheme and nomenclature for the scorpion homeobox genes was primarily based on the published descriptions (Holland et al., 2007). All of the identified homeobox genes encode proteins containing complete or partial homeodomain sequences. The pseudogenes were not analyzed. Furthermore, we performed a scaffold location analysis on homeobox genes in the scorpion *M. martensii* genome to investigate the Hox clusters.

2.2. Sequence alignment and phylogenetic analysis

The 156 homeodomain sequences from 149 homeobox genes resulting from BLASTP analysis of HomeoDB were aligned by ClustalX 1.83. The alignment file was produced by DNAMAN (highlight homology level $\geq 33\%$). A phylogenetic tree of the scorpion homeodomain family was reconstructed using the maximum likelihood (ML) algorithms implemented in Phyml 3.0 with the 100-fold bootstrap test after the best models were estimated by the ProtTest 2.4 server (http://darwin.uvigo.es/software/prottest2_server.html). The tree diagram was generated and edited using FigTree (Version 1.40) and Mega (5.1).

2.3. Gene expression analysis

To determine the cDNA/ESTs available for each identified homeobox gene, a local TBLASTN search was performed on transcriptome data from the scorpion *M. martensii* using the 149 putative scorpion homeobox genes as queries. Those consistent in the reciprocal TBLAST searches (identity value of homeodomain regions > 95%) were considered cDNA/EST evidence.

3. Results and discussion

3.1. Identification and classification of homeobox genes from the scorpion genome

We identified 149 homeobox genes from the scorpion genome, which represents *M. martensii* homeobox genes (Fig. 1, and Supplementary files 2&3). This is the first survey of scorpion Hox genes. All of the identified homeobox genes were distributed on 134 contigs (Supplementary file 4). The gene expression of the identified scorpion homeobox genes was analyzed using whole transcriptome data from the scorpion *M. martensii*. Of the 149 putative scorpion homeobox genes, 121 have EST evidence. More than 80% (121/149) of the putative homeobox genes are transcribed in the adult scorpion (Supplementary file 5).

Using the recently reported classification scheme as a reference (Holland et al., 2007; Chai et al., 2008), we classified 143 of the 149 identified scorpion homeobox genes into eight classes that cover 82 families. The remaining six scorpion homeobox genes, MMA08570, MMA03106, MMA18441, MMA32551, MMA11725 and MMA49726, could not be assigned to a corresponding family based on the current classification scheme. Phylogenetic relationships were analyzed between the scorpion homeobox genes and their homologs from beetle *T. castaneum*, honeybee *A. mellifera*, and human *H. sapiens* genomes. Most of the identified scorpion homeobox genes could form a monophyletic group with their homologs supported

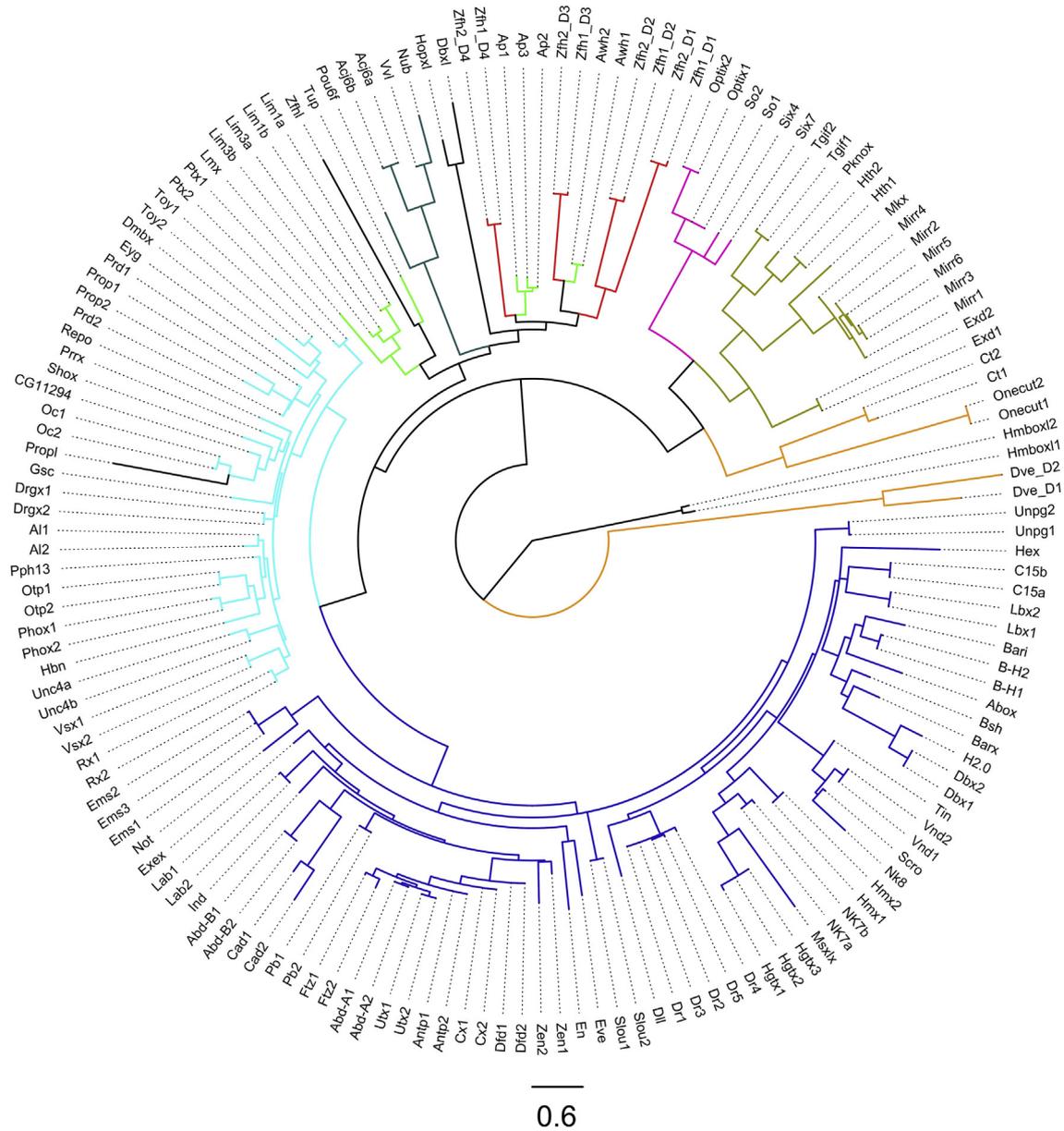


Fig. 1. Diversity of homeobox members from the scorpion *M. martensii*. Phylogenetic analysis of all scorpion homeodomains was performed using the maximum likelihood (ML) algorithms implemented by Phyml 3.0. Several classes are not recovered as monophyletic groups. Due to the short length of the sequence alignment and the complex modes of evolution, this tree should not be used to infer accurate evolutionary history, but it is presented to demonstrate the diversity of the scorpion homeodomain sequences.

by a high confidence bootstrap value. Detailed information about the members of each class is given below (Supplementary file 2).

3.1.1. The ANTP homeobox class

The ANTP class was named after the *Drosophila* homeotic gene *Antennapedia*, which was further divided into two subclasses, the HOXL subclass and NKL subclass (the contents of these subclasses followed HomeoDB. Sixty-seven homeobox genes were identified as belonging to the ANTP class in the scorpion *M. martensii* and were categorized into 34 families (Figs. 2 & 3, and Supplementary file 2). The HOXL subclass of *M. martensii* was composed of 27 genes belonging to 11 families: Cdx (2), Evx (1), Gbx (3), Hox1 (2), Hox2 (2), Hox3 (2), Hox4 (2), Hox5 (2), Hox6–8 (8), Hox9–13 (2) and Mnx (1). All of these families are non-insect specific or vertebrate specific. Members of this subclass could form a monophyletic group with their homologs from other genomes (Fig. 2). The NKL subclass of *M. martensii* includes 40 genes belonging to 24 families:

Abox (1), Barhl (2), Bari (1), Barx (1), Bsx (1), Dbx (2), Dlx (1), Emx (3), En (1), Hhex (1), Hlx (1), Lbx (2), Msx (5), Msxlx (1), Nk1 (2), Nk2.1 (2), NK2.2 (2), NK 4(1), NK5 (2), NK6 (2), NK7 (2), NK8 (1), Noto (1) and Tlx (2). Among the NKL subclass, four families, Abox, Bari, Msxlx and NK7, are found in insects, mollusks and nematodes. Members of this subclass from the scorpion *M. martensii* genome form a monophyletic group with their homologs from other genomes (Fig. 3). In total, the number of scorpion ANTP class genes closely approximates those of chicken and amphioxus and is greater than those of insects and nematode but less than those of human, mouse, zebrafish and frog (Supplementary file 1).

3.1.2. The PRD homeobox class

The PRD class was named after the *D. melanogaster* *paired* gene, and includes most of the Pax genes and many non-Pax gene families. Thirty-three homeobox genes were identified as belonging to the PRD class, covering 19 families (Supplementary file 6 (A)).

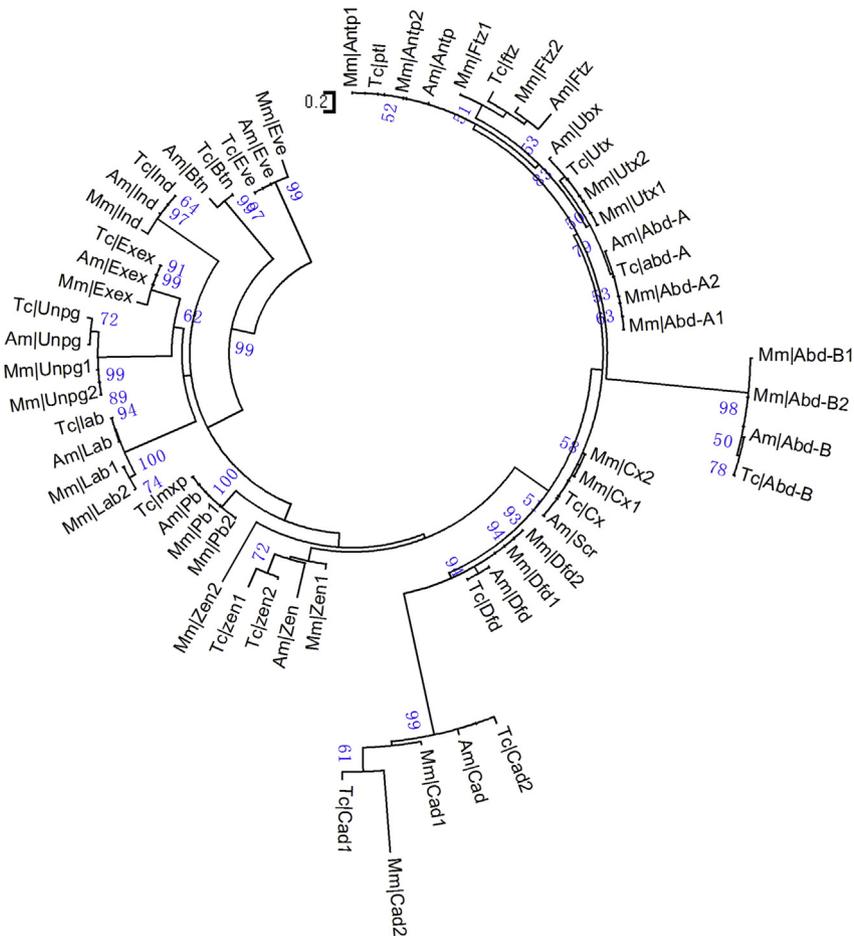


Fig. 2. Maximum likelihood phylogenetic tree of the HOXL subclass (belonging to ANTP class) homeodomains from scorpion, beetle and honeybee. The maximum likelihood tree was generated using Phym1 (version 3.0) and rooted with Eve genes (Evx family).

Almost all members of this class formed a monophyletic group with their homologs from other animal groups (such as beetle and human). The PRD class is further divided into two subclasses, PAX and PAXL, based on whether it possessed a conserved paired-box (Chai et al., 2008). We found five members belonging to the PAX subclass in the scorpion genome. All 5 members are categorized into 2 families: Pax3/7 (2) and Pax4/6 (3). The PAX genes did not segregate into a single group in the phylogenetic tree. The PAXL subclass contains 17 families in the scorpion *M. martensii*, including Arx (3), CG11294 (1), Dmbx (1), Drgx (2), Gsc (1), Hbn (1), Otp (2), Otx (2), Phox (2), Pitx (2), Prop (2), Prrx (1), Rax (1), Repo (1), Shox (1), Uncx (2) and Vsx (2). Among these 17 families, three families (CG11294, Hbn, and Repo) are found in the scorpion *M. martensii* and some insects. Amphioxus, *B. floridae* has Repo (Takatori et al., 2008). Because no corresponding vertebrate homologous genes were identified. The result implies that these families may play an important role in the development of invertebrates. In total, the number of scorpion PRD class genes approximates those of chicken, frog, amphioxus and fruit fly and is greater than those of the beetle, honeybee and nematode but less than those of the human, mouse and zebrafish (Supplementary file 1).

3.1.3. The LIM homeobox class

The LIM homeobox class genes encode proteins with two LIM domains and a homeodomain with several diagnostic amino acid residues (Holland et al., 2007). We found a total of 11 homeobox

genes belonging to the LIM class in the scorpion genome. All of these genes formed a monophyletic group with their homologs from beetle and human in a family-group manner (Supplementary file 6 (B)). These 11 homeobox genes are categorized into 6 families: Isl (1), Lhx1/5 (2), Lhx2/9 (3), Lhx3/4(2), Lhx6/8 (2), and Lmx (1). Most of the genes identified in this class encode proteins containing both the homeodomain and one or more LIM domains, and the few without LIM domains are putatively partial genes. The number of LIM class genes in the scorpion *M. martensii* approximates those of human, mouse, chicken and frog and is greater than those of amphioxus, fruit fly, beetle, honeybee and nematode but less than that of zebrafish (Supplementary file 1). None of these organisms have more than 20 genes in this class.

3.1.4. The POU homeobox class

The POU class proteins possess a 60 amino acid homeodomain and 75 amino acid N-terminal POU domain. Five POU homeobox class candidate genes were discovered in the scorpion genome and belong to four families (Pou2, Pou3, Pou4 and Pou6). As shown in Supplementary file 6 (B), each member formed a monophyletic group with their beetle and human homologs. The number of members in the POU class from the scorpion genome is the same as that of fruit fly, similar to those of chicken, amphioxus, beetle, honeybee and nematode, and less than those of human, mouse, frog and zebrafish (Supplementary file 1). In these animals, there are no more than 20 genes in this class.

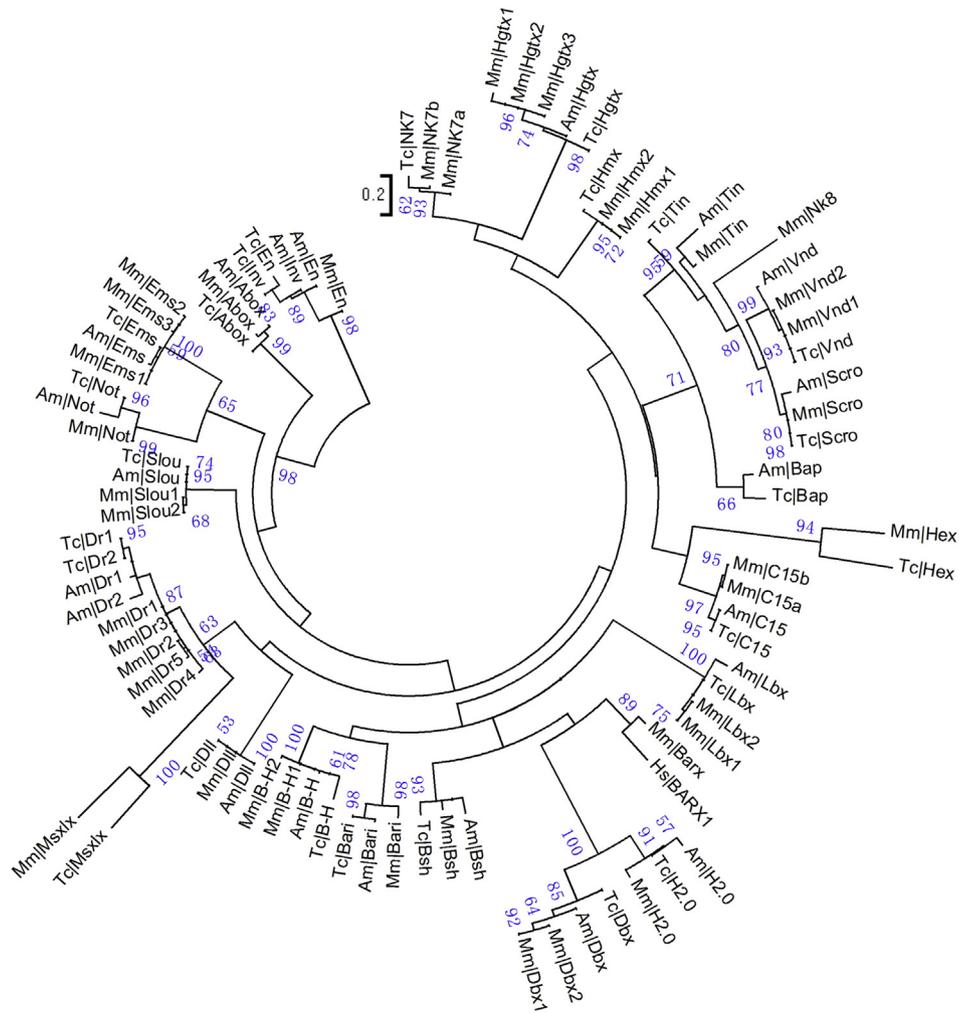


Fig. 3. Maximum likelihood phylogenetic tree of the NKL subclass (belonging to ANTP class) homeodomains from scorpion, beetle and honeybee. The maximum likelihood tree was generated using PhymI (version 3.0) and rooted with Inv/En genes (En family).

3.1.5. The SINE homeobox class

We identified six SINE class homeobox genes in the scorpion *M. martensii* genome. These six genes were further categorized into four families, including three known families (Six1/2, Six3/6 and Six4/5) and one new family (Six 7). The Six1/2 and Six3/6 families each have two members, and the Six4/5 family has one member in *M. martensii*. As shown in [Supplementary file 6 \(B\)](#), each member of Six1/2, Six3/6 and Six4/5 families formed a monophyletic group with their homologs from beetle and human. We named it as the single member of a new family: Six 7. In total, the diversity of the SINE class homeobox genes in the scorpion genome is the same as those of human, mouse and frog, similar to that of nematode, greater than those of chicken, amphioxus, fruit fly, beetle and honeybee, and less than that of zebrafish ([Supplementary file 1](#)).

3.1.6. The CUT homeobox class

The CUT class consists of three families (Onecut, Cux and Cmp) in invertebrates ([Chai et al., 2008](#)). Onecut and Cux family members are reported to encode proteins with a homeodomain and one or more 75 amino acid DNA-binding CUT domains, while members of the Cmp family encode proteins lacking the CUT domain ([Chai et al., 2008](#)). Three members of the CUT class were previously identified in the silkworm genome and each member belonged to one family: Cmp (1), Cux (1) and Onecut (1) ([Chai et al., 2008](#)). The CUT gene

diversity in amphioxus and sea urchins was recently analyzed ([Takatori et al., 2008](#); [Cameron et al., 2006](#)). This study reveals five members of the CUT class in the scorpion *M. martensii* genome: Cmp (1), Cux (2) and Onecut (2). As shown in [Supplementary file 6 \(B\)](#), members of the Onecut and Cux families grouped together, but the Cmp family did not, which is consistent with previously reported results ([Holland et al., 2007](#); [Chai et al., 2008](#)). Due to their partial sequences, we only found the homeodomain for these genes and did not find their CUT domains. Each member formed a monophyletic group with their homologs from beetle and human. The number of the CUT class genes in the scorpion *M. martensii* is similar to those of human, mouse, chicken, frog, amphioxus and nematode, greater than those of fruit fly, beetle and honeybee, and less than that of zebrafish ([Supplementary file 1](#)). Similarly, all of the analyzed animals have no more than 10 genes in this class.

3.1.7. The TALE homeobox class

The TALE class includes six gene families: Irx, Meis, Mlx, Pbx, Pknox, and Tgif ([Holland et al., 2007](#)). Several of these gene families also have conserved domains outside the homeodomain, including the IRO domain, MEIS domain, and MKX domains ([Holland et al., 2007](#)). This study shows that 14 scorpion homeobox genes belong to the TALE class, which covers all six gene families of this class. The Irx family of *M. martensii* consists of six members, and the three

other families (Meis, Pbx, and Tgif) have two members each. The Mxk and Pknox families have only one member each (Supplementary file 2). Three members of the Irx family, MMa19657, MMa19660 and MMa19666, are tandemly arranged in Contig317529, while the other three members of this family, MMa50880, MMa50881, and MMa50884, are similarly tandemly arranged in Contig350484 (Supplementary file 4). All of the 6 mirror genes encode proteins containing both the homeodomain and one IRO domain (Supplementary file 2). As shown in Supplementary file 6 (B), members of the TALE class formed a relatively independent group in the phylogenetic tree. In total, the gene diversity of the TALE class in the scorpion genome is similar to those of chicken and frog, greater than those of amphioxus, fruit fly, beetle, honeybee and nematode, and less than that of human, mouse, and zebrafish (Supplementary file 1).

3.1.8. The ZFH homeobox class

Genes in the zinc finger (ZF) class, also called the ZFH class, encode proteins with homeodomain and zinc finger (Znf) motifs. The ZF class includes five families: Adnp, Zfhx, Zeb, Zhx/Homez, and

Tshz (Holland et al., 2007). The scorpion *M. martensii* genome has two members of the ZFH class, and the number is similar to those of the fruit fly *D. melanogaster* and the silkworm *B. mori*. Both members belong to the Zfhx family: one encodes a protein containing 32 Znf domains and four homeodomains, and the other encodes a protein containing 14 Znf domains and four homeodomains. As previously reported in *B. mori* and *B. floridae* (Chai et al., 2008; Takatori et al., 2008), phylogenetic analyses of scorpion homeo-domain sequences from the ZFH class did not form a monophyletic group (Supplementary file 6 (B)). In total, the scorpion *M. martensii* has an identical number of members (2) in the ZFH class, which is the same as those for fruit fly, beetle, and nematode. The honeybee has only one member identified, and the amphioxus, human, mouse, chicken, frog, and zebrafish have 5, 14, 14, 11, 14, and 17, respectively (Supplementary file 1). The ZFH class has a diverse number of members in animals.

3.1.9. The other homeobox classes

We did not find any members of the CERS, HNF, and PROS classes in the scorpion genome, whereas they have been

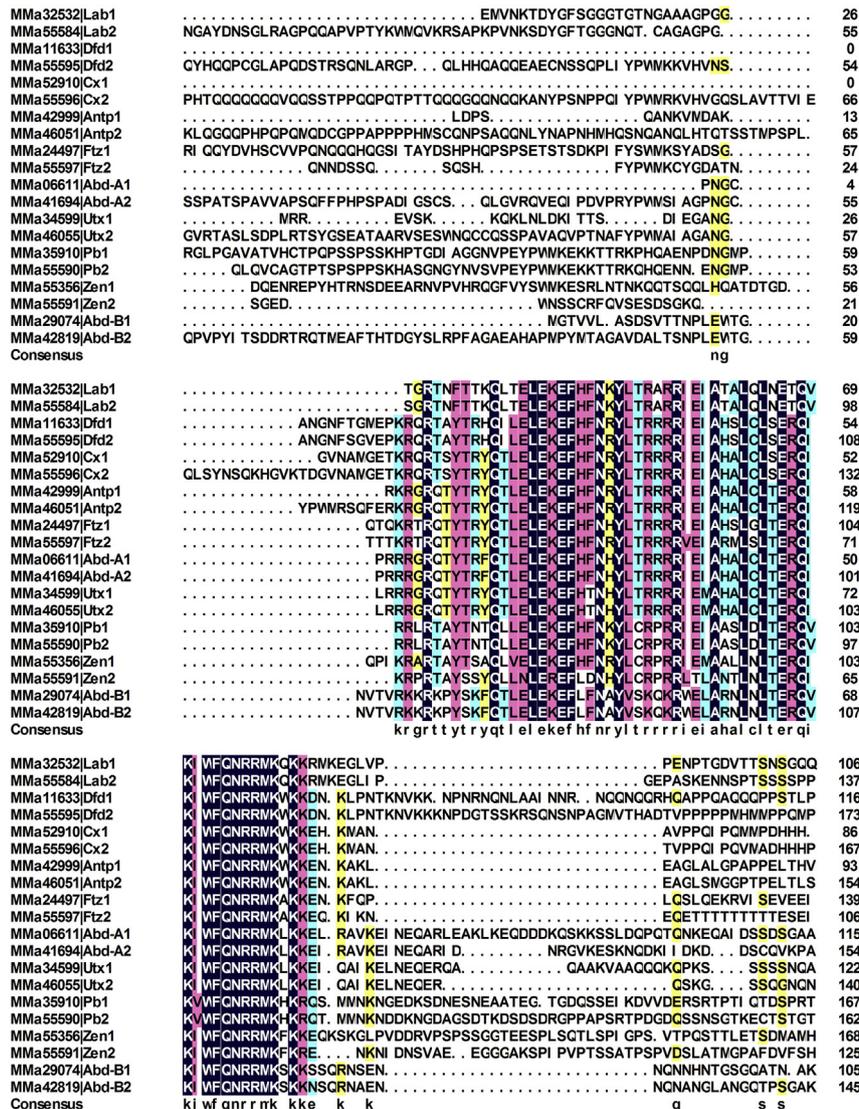


Fig. 4. Multiple sequence alignments of 20 Hox cluster members from the genome of the scorpion *M. martensii*. Each of ten Hox cluster members has two copies. Each Hox cluster member pair shares high homology, but they are also different distinctly.

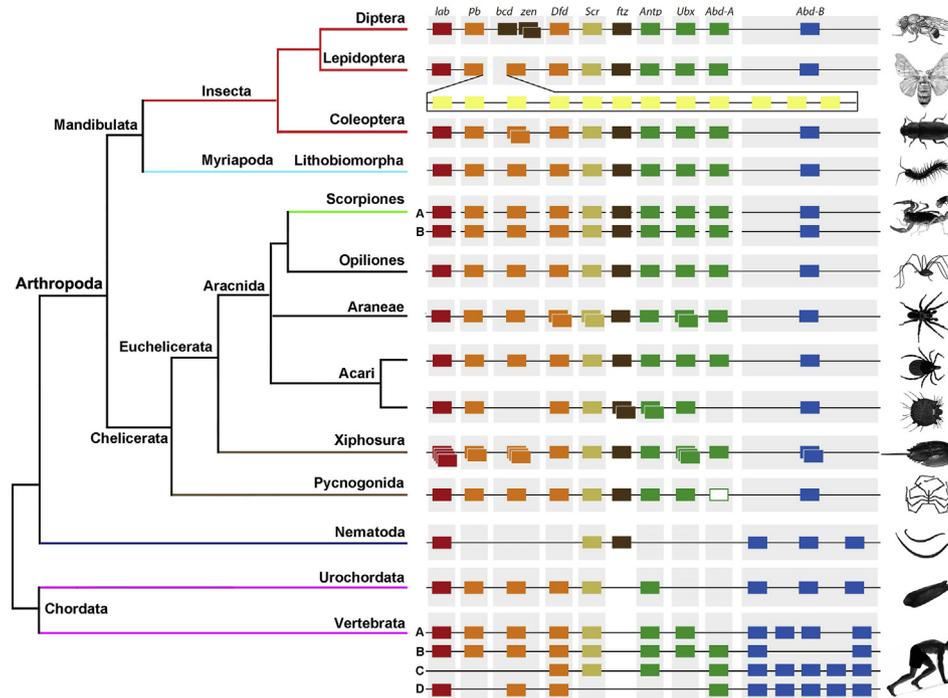


Fig. 6. Duplication and divergence of Hox cluster genes from the different animal groups (Damen et al., 1998; Abzhanov et al., 1999; Jager et al., 2006; Manuel et al., 2006; Schwager et al., 2007; Chai et al., 2008; Damen et al., 2010; Sharma et al., 2012). The gene contents were followed the references and not found throughout these clades. Horizontal line is not the evidence for clustering along the chromosome. We hypothesized that there are two Hox clusters in scorpions, which is still confirmed in future work. The tree represents the generally accepted arthropod relationships, as previously described (Regier et al., 2010; Shultz et al., 2007).

et al., 1997), sea urchins (Cameron et al., 2006) and chelicerates (Damen, 2010). Vertebrates usually have more Hox clusters, e.g., the two, three or four Hox clusters found in shark (chondrichthyes) (Kim et al., 2000; Oulion et al., 2010), the four in sarcopterygian vertebrates (HoxA–D) (Liang et al., 2011) and the seven or eight in teleosts (Pascual-Anaya et al., 2012; Crow et al., 2006; Hoegg et al., 2007). Most of these clusters are not complete.

There are some examples of the duplication and reduction of Hox genes in chelicerates. Twenty-eight different small homeobox gene fragments were found in the horseshoe crab *Limulus polyphemus* (Cartwright et al., 1993). Two copies of the *Deformed* (*Dfd*) gene were identified in *Achaearanea* (spider) (Abzhanov et al., 1999), and a duplicated *Ultrabithorax* (*Ubx*) gene was described in *Cupiennius* (spider) (Damen et al., 1998). Nine Hox genes without the *abdA* gene were recovered in *Endeis spinosa*, and a derived sequence for an *abdA* gene was found in the sea spider *Nymphon gracile* (Damen et al., 2010; Manuel et al., 2006). At least three Hox genes (*Ultrabithorax*, *Deformed* and *Sex comb reduced*) are present as two copies in the spider *Cupiennius salei* (Schwager et al., 2007). The spider mite, *T. urticae*, contains an eight-gene cluster of the canonical 10 genes with the *ftz* gene present in a duplicated formation, but both the *Hox3* and abdominal A (*abdA*) genes were lost (Grbić et al., 2011).

Not only were all ten Hox cluster genes identified in the scorpion *M. martensii*, but also each of the ten Hox cluster members has two copies. Each Hox cluster member pair shares high similarity, but they are also different. Importantly, the genomic sequence of scaffold Contig352199 revealed five different Hox cluster genes that are closely linked to each other. Our data imply a hypothesis that the scorpion *M. martensii* has two complete Hox clusters (Fig. 6) in its genome. It is difficult to determine the functional difference between the two putative Hox clusters in *M. martensii*, but their sequence differences and transcription pattern suggests that the

members of the two putative Hox clusters play roles in scorpion development, which may be associated with their unique morphological structures.

As shown in Fig. 6, although the duplication and deletion of Hox cluster genes could be confirmed in some groups, they was reported to have only one Hox cluster (or broken apart), like as spider mites, ticks and the Mandibulata arthropods. Scorpions have not just one or more Hox cluster genes reproduced in situ, but may have two Hox clusters, which may be the first example of integrally duplicated Hox clusters. Interestingly and coincidentally, researchers hypothesize that *Cupiennius salei* (spider) may have two Hox clusters (Schwager et al., 2007).

In summary, as the first complete genome-wide analysis of homeobox genes in Chelicerata, our work identified a complete set of homeobox genes in the scorpion *M. martensii* (149 homeobox genes belonging to eight classes covering 82 families). Combined with the results of Sharma et al. (2014) that the scorpion *C. sculpturatus* has two paralogues of all Hox genes except Hox3, the genome-wide analysis of homeobox genes from *M. martensii* suggested that the scorpion *M. martensii* may have two Hox clusters on its genome, which is a new evidence for Hox cluster and/or whole genome duplication in scorpions. Our work not only reveals the repertoire of scorpion, arachnid and chelicerate homeobox genes but also elucidates the evolution of arthropod homeobox genes.

Acknowledgments

This work was supported by the grants from National Science Fund for Excellent Young Scholars (No. 31422049), Hubei Science Fund for Excellent Scholars (No. ZRZ2015000090), Fundamental Research Fund for the Central Universities in China (No. 2042014kf0205), and China Scholarship Council (No. 201308420306).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.ibmb.2015.04.002>.

References

- Abzhanov, A., Popadic, A., Kaufman, T.C., 1999. Chelicerate Hox genes and the homology of arthropod segments. *Evol. Dev.* 1, 77–89.
- Bürglin, T.R., 1994. A comprehensive classification of homeobox genes. In: Duboule, D. (Ed.), *Guidebook to the Homeobox Genes*. Oxford University Press, New York, pp. 25–71.
- Bürglin, T.R., 2005. Homeodomain proteins. In: Meyers, R.A. (Ed.), *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, second ed., vol. 6. Wiley-VCH Verlag GmbH & Co, Weinheim, pp. 179–222.
- Cameron, R.A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J.P., Berneym, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P.M., Davidson, E.H., Peterson, K.J., Hood, L., 2006. Unusual gene order and organization of the sea urchin Hox cluster. *J. Exp. Zool. B. Mol. Dev. Evol.* 306, 45–58.
- Cao, Z.J., Yu, Y., Wu, Y.L., Hao, P., Di, Z.Y., He, Y.W., Chen, Z.Y., Yang, W.S., Shen, Z.Y., He, X.H., Sheng, J., Xu, X.B., Pan, B.H., Feng, J., Yang, X.J., Hong, W., Zhao, W.J., Li, Z.J., Huang, K., Li, T., Kong, Y.M., Liu, H., Jiang, D.H., Zhang, B.Y., Hu, J., Hu, Y.T., Wang, B., Dai, J.L., Yuan, B.F., Feng, Y.Q., Huang, W., Xing, X.J., Zhao, G.P., Li, X., Li, Y.X., Li, W.X., 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropod. *Nat. Commun.* 3602, 1–10.
- Cartwright, P., Dick, M., Buss, L.W., 1993. HOM/Hox type homeoboxes in the chelicerate *Limulus polyphemus*. *Mol. Phyl. Evol.* 2, 185–192.
- Chai, C.L., Zhang, Z., Huang, F.F., Wang, X.Y., Yu, Q.Y., Liu, B.B., Tian, T., Xia, Q.Y., Lu, C., Xiang, Z.H., 2008. A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 38, 1111–1120.
- Crow, K.D., Stadler, P.F., Lynch, V.J., Amemiya, C., Wagner, G.P., 2006. The “fish-specific” Hox cluster duplication is coincident with the origin of teleosts. *Mol. Biol. Evol.* 23, 121–136.
- Damen, W.G., 2010. Hox genes and the body plans of chelicerates and pycnogonids. *Adv. Exp. Med. Biol.* 689, 125–132.
- Damen, W.G., Hausdorf, M., Seyfarth, E.A., Tautz, D., 1998. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10665–10670.
- Dhawan, S., Gopinathan, K.P., 2005. Phylogeny of the insect homeobox gene (Hox) cluster. *Genomics Proteomics Bioinform.* 3, 42–46.
- Frobus, A.C., Matus, D.Q., Seaver, E.C., 2008. Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan *Capitella* sp. I. *PLoS One* 3, e4004.
- Grbić, M., Van Leeuwen, T., Clark, R.M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E.J., Dermauw, W., Ngoc, P.C., Ortego, F., Hernández-Crespo, P., Díaz, I., Martínez, M., Navajas, M., Sucena, É., Magalhães, S., Nagy, L., Pace, R.M., Djuranović, S., Smaghe, G., Iga, M., Christiaens, O., Veenstra, J.A., Ewer, J., Villalobos, R.M., Hutter, J.L., Hudson, S.D., Velez, M., Yi, S.V., Zeng, J., Pires-daSilva, A., Roch, F., Cazaux, M., Navarro, M., Zhurov, V., Acevedo, G., Bjelica, A., Fawcett, J.A., Bonnet, E., Martens, C., Baele, G., Wissler, L., Sanchez-Rodriguez, A., Tirry, L., Blais, C., Demeestere, K., Henz, S.R., Gregory, T.R., Mathieu, J., Verdon, L., Farinelli, L., Schmutz, J., Lindquist, E., Feyereisen, R., Van de Peer, Y., 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487–492.
- Grenier, J.K., Garber, T.L., Warren, R., Whittington, P.M., Carroll, S., 1997. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr. Biol.* 7, 547–553.
- Heffer, A., Pick, L., 2013. Conservation and variation in hox genes: how insect models pioneered the evo-devo field. *Annu. Rev. Entomol.* 58, 161–179.
- Hoegg, S., Boore, J.L., Kuehl, J.V., Meyer, A., 2007. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 8, 317.
- Holland, P.W., Booth, H.A., Bruford, E.A., 2007. Classification and nomenclature of all human homeobox genes. *BMC Biol.* 5, 47.
- Irvine, S.Q., Warinner, S.A., Hunter, J.D., Martindale, M.Q., 1997. A survey of homeobox genes in *Chaetopterus variopedatus* and analysis of polychaete homeodomain. *Mol. Phylogenet. Evol.* 7, 331–345.
- Jager, M., Murienne, J., Clabaut, C., Deutsch, J., Le Guyader, H., Manuel, M., 2006. Homology of arthropod anterior appendages revealed by Hox gene expression in a sea spider. *Nature* 441, 506–508.
- Kim, C.B., Amemiya, C., Bailey, W., Kawasaki, K., Mezey, J., Miller, W., Minoshima, S., Shimizu, N., Wagner, G., Ruddle, F., 2000. Hox cluster genomics in the horn shark, *Heterodontus francisci*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1655–1660.
- Liang, D., Wu, R., Geng, J., Wang, C., Zhang, P., 2011. A general scenario of Hox gene inventory variation among major sarcopterygian lineages. *BMC Evol. Biol.* 11, 25.
- Manuel, M., Jager, M., Murienne, J., Clabaut, C., Le Guyader, H., 2006. Hox genes in sea spiders (*Pycnogonida*) and the homology of arthropod head segments. *Dev. Genes. Evol.* 216, 481–491.
- McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A., Gehring, W.J., 1984. A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* 308, 428–433.
- Oulion, S., Debais-Thibaud, M., d'Aubenton-Carafa, Y., Thermes, C., Da Silva, C., Bernard-Samain, S., Gavory, F., Wincker, P., Mazan, S., Casane, D., 2010. Evolution of Hox gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. *Mol. Biol. Evol.* 27, 2829–2838.
- Pascual-Anaya, J., Adachi, N., Alvarez, S., Kuratani, S., D'Aniello, S., Garcia-Fernández, J., 2012. Broken colinearity of the amphioxus Hox cluster. *Evodevo* 3, 28.
- Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463, 1079–1083.
- Schwager, E.E., Schoppmeier, M., Pechmann, M., Damen, W.G., 2007. Duplicated Hox genes in the spider *Cupiennius salei*. *Front. Zool.* 4, 10.
- Scott, M.P., Weiner, A.J., 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 81, 4115–4119.
- Sharma, P.P., Schwager, E.E., Extavour, C.G., Giribet, G., 2012. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.
- Sharma, P.P., Schwager, E.E., Extavour, C.G., Wheeler, W.C., 2014. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281, 20140661.
- Shultz, J.W., 2007. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zool. J. Linn. Soc.* 150, 221–265.
- Takatori, N., Butts, T., Candiani, S., Pestarino, M., Ferrier, D.E., Saiga, H., Holland, P.W., 2008. Comprehensive survey and classification of homeobox genes in the genome of amphioxus, *Branchiostoma floridae*. *Dev. Genes. Evol.* 218, 579–590.
- Zhong, Y.F., Holland, P.W., 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* 13, 567–568.